

Fig. 1: Peter and the wolf: Score

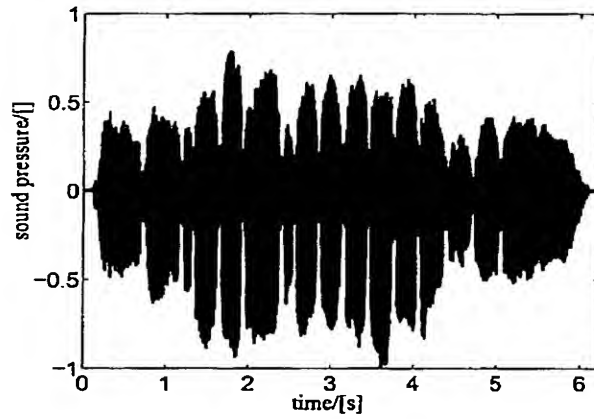


Fig. 2: Peter and the wolf: Sound wave

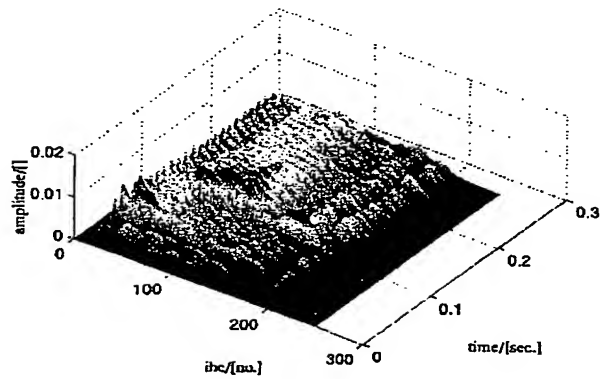


Fig. 3: IHC cleft contents

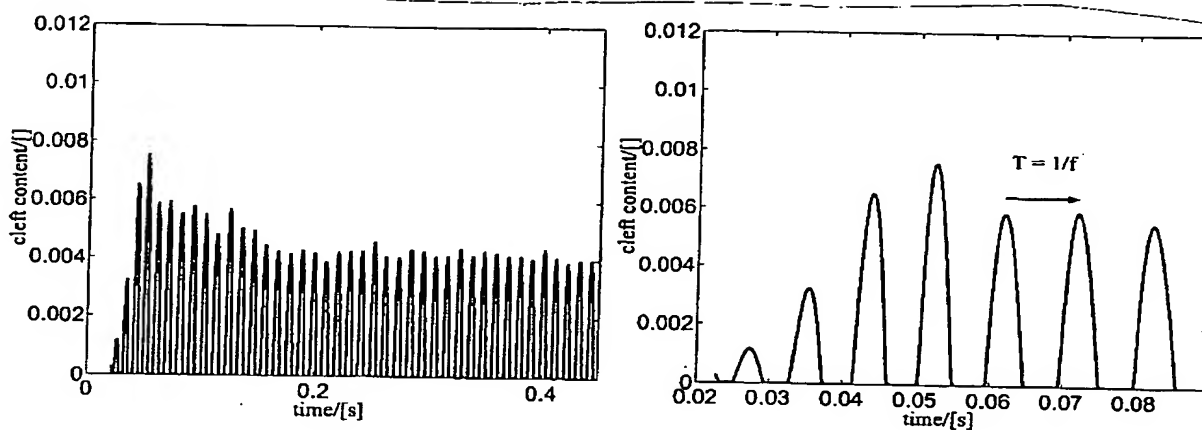


Fig. 4: Cleft content of IHC #12

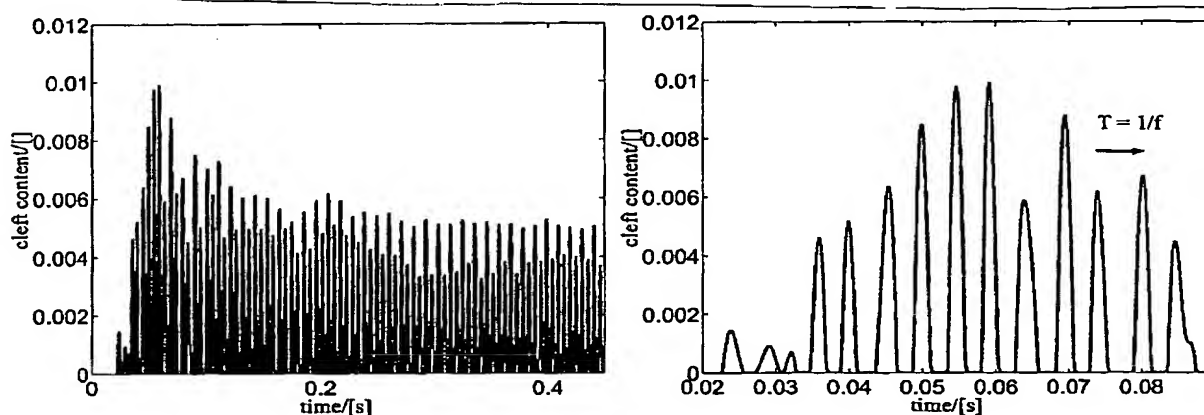


Fig. 5: Cleft content of IHC #25

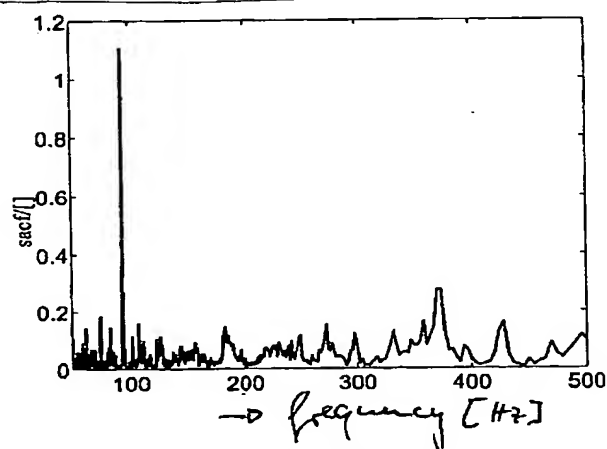


Fig. 6: SACF histogram

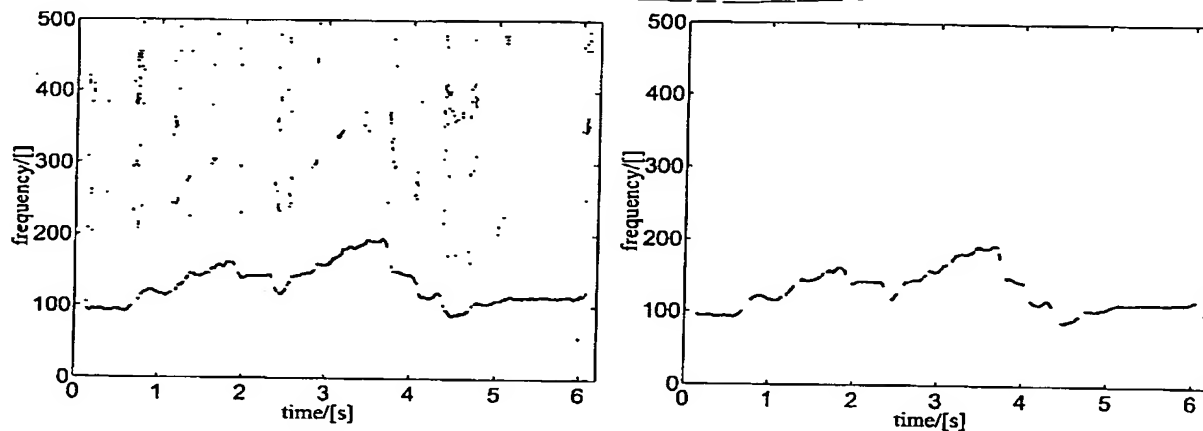


Fig. 7: Initial pitch estimation vs. cleaned pitch trajectories

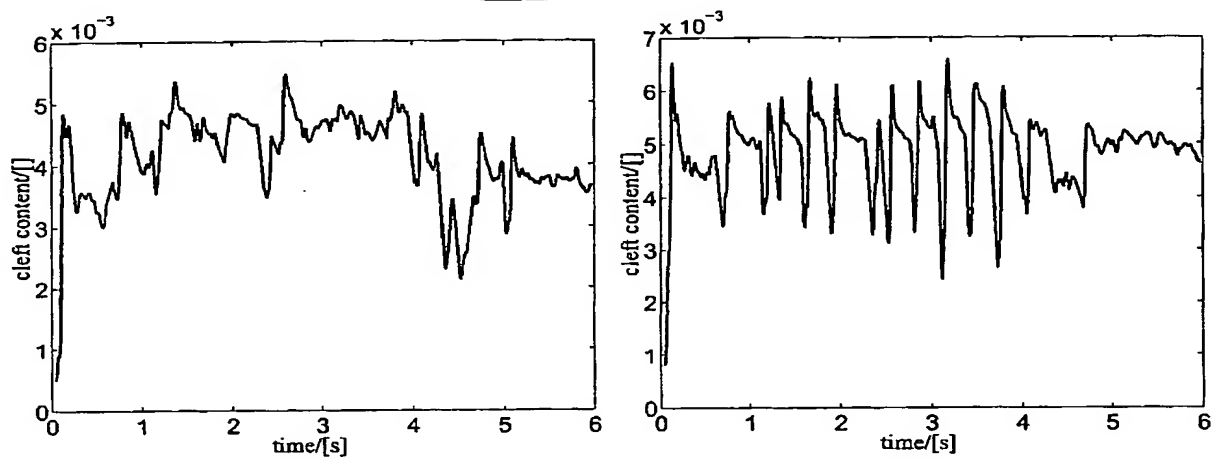


Fig. 8: Envelopes of transmitter substance: 1st and 4th partial

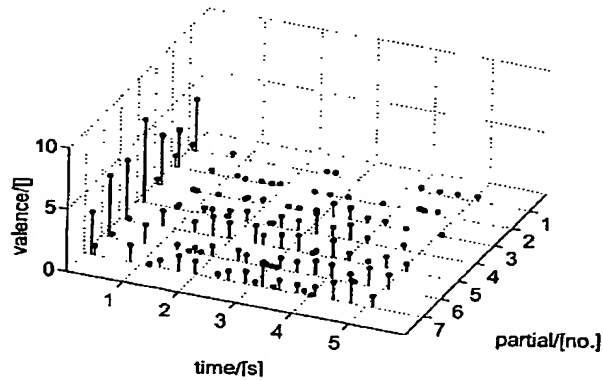
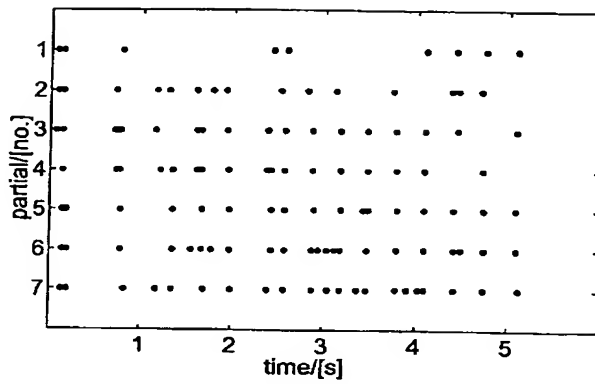


Fig. 9: Onset map

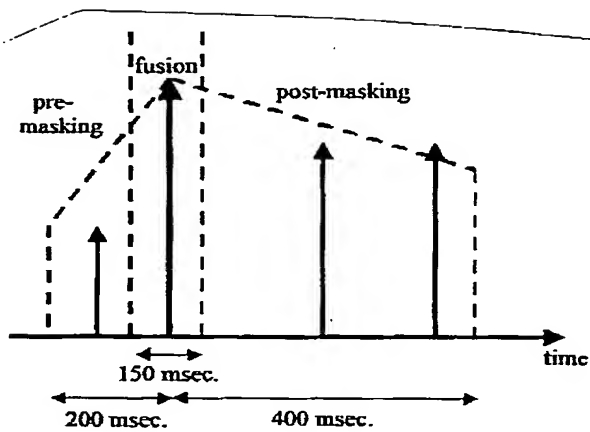


Fig. 10: Onset masking

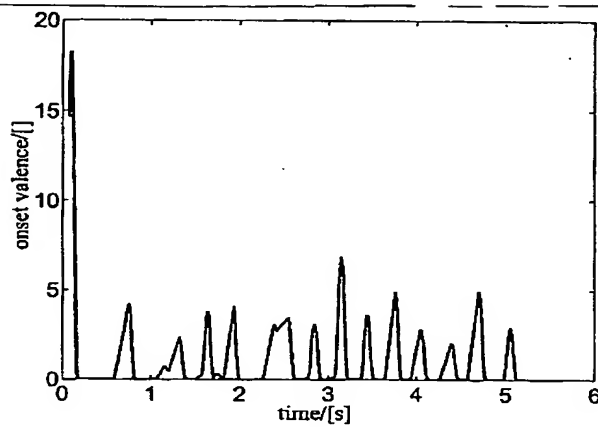


Fig. 11: Onset histogram

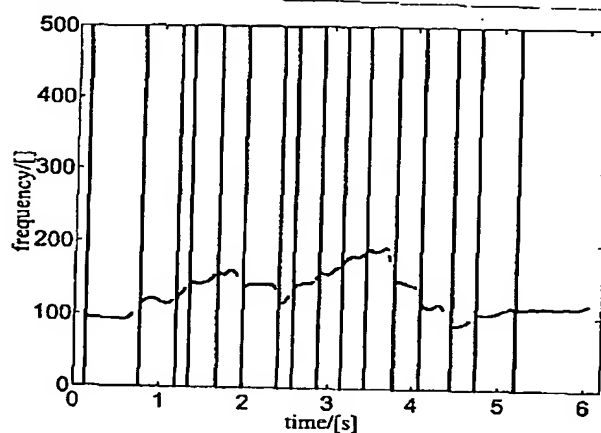


Fig. 12: Pitch trajectories, segmented

partial [no.]	onset time [s]	frequency [Hz]	amplitude []	IHC peak [no.]	IHC resonance [no.]
1	0.187	236.4	0.00878	27	30
2	0.218	467.8	0.00555	48	14
3	0.192	703.5	0.00862	73	36
4	0.219	878.2	0.00735	94	2
5	0.201	1170.6	0.00733	102	12
6	0.202	1405.3	0.00776	113	12
7	0.206	1640.8	0.00783	123	16

Fig. 13: Feature values, clarinet Bb3

		EarAnalyzer	Extreme	Hough
database 1 (200 pieces)	Position 1	82.8 %	59.5 %	39.9 %
	Top 10	92.5 %	75.1 %	56.9 %
database 2 (1024 pieces)	Position 1	78.5 %	53.5 %	32.0 %
	Top 10	88.9 %	67.9 %	42.9 %

Fig. 14: Query-By-Humming: 1152 query inputs

		EarAnalyzer	Extreme	Hough
Fullrate	Position 1	79.3 %	49.4 %	42.0 %
	Top 10	88.9 %	67.1 %	56.3 %
Enhanced Fullrate	Position 1	80.6 %	60.4 %	45.0 %
	Top 10	91.2 %	73.9 %	62.7 %
Half-rate	Position 1	69.9 %	48.4 %	30.2 %
	Top 10	82.5 %	65.4 %	49.0 %
Original	Position 1	82.8 %	59.5 %	39.9 %
	Top 10	92.5 %	75.1 %	56.9 %

Fig. 15: Query-By-Humming incl. GSM distortion: 1152 query inputs

training vs. query	McGill [%]	Gdansk [%]	Fraunhofer [%]
McGill	100 (100 100 100)	80 (94 47 94)	88 (97 67 97)
Gdansk	85 (75 87 94)	100 (100 100 100)	87 (86 73 100)
Fraunhofer	81 (58 93 94)	72 (64 60 94)	100 (100 100 100)
McGill + Gdansk	99 (97 100 100)	100 (100 100 100)	91 (92 83 97)
McGill + Fraunhofer	100 (100 100 100)	83 (92 53 100)	100 (100 100 100)
Gdansk + Fraunhofer	88 (78 93 94)	100 (100 100 100)	100 (100 100 100)

Fig. 16 : Instrument recognition rates [total (clarinet|oboe|bassoon)]

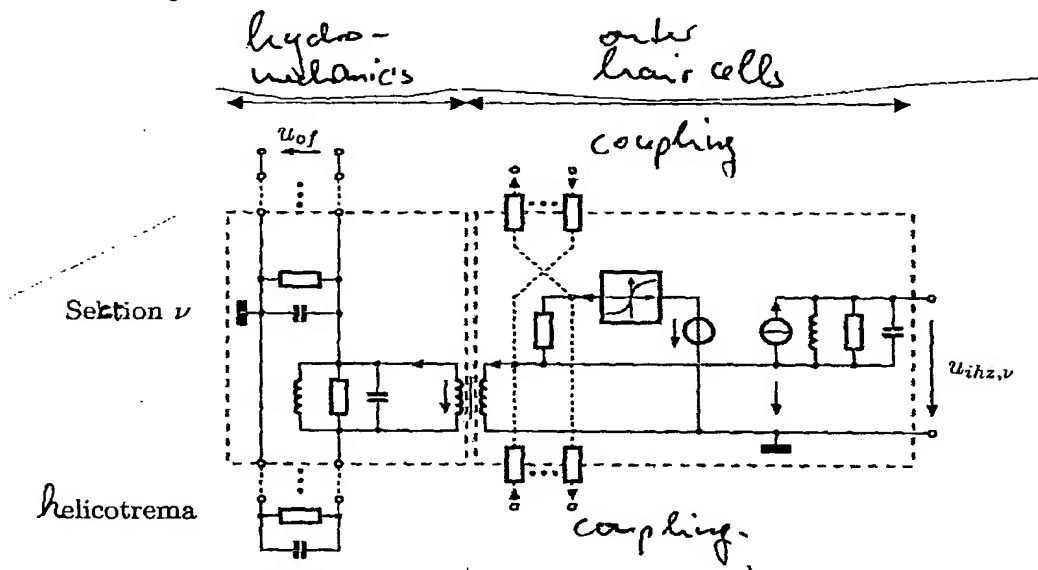


Fig. 17 : extended analog model by Baningarte

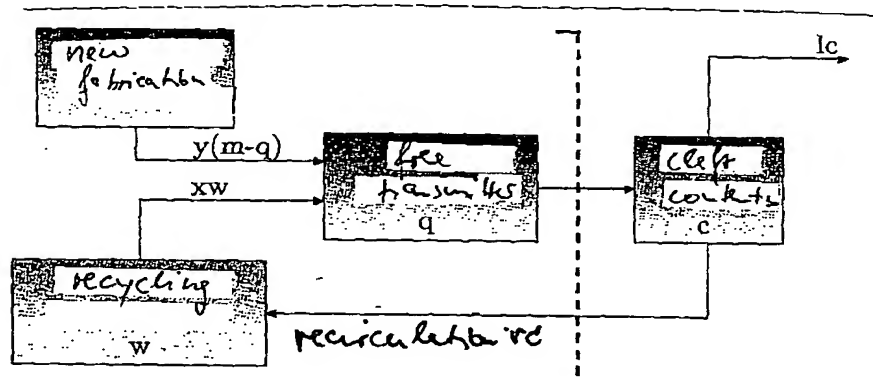


Fig. 18: hair cell model by Meddis

$$k = \begin{cases} gdt \cdot \left(\frac{s+A}{s+A+B} \right) & : s+A \geq 0 \\ 0 & : s+A < 0. \end{cases}$$

$$\frac{dq}{dt} = y(m-q) + xw - kq,$$

$$\frac{dc}{dt} = kq - lc - rc,$$

$$\frac{dw}{dt} = rc - xw.$$

A	=	10
B	=	3000
x	=	66.31
g	=	1000
y	=	5.05
m	=	1
l	=	2500
r	=	6580

Fig. 19: mathematical description of the model in Fig. 18

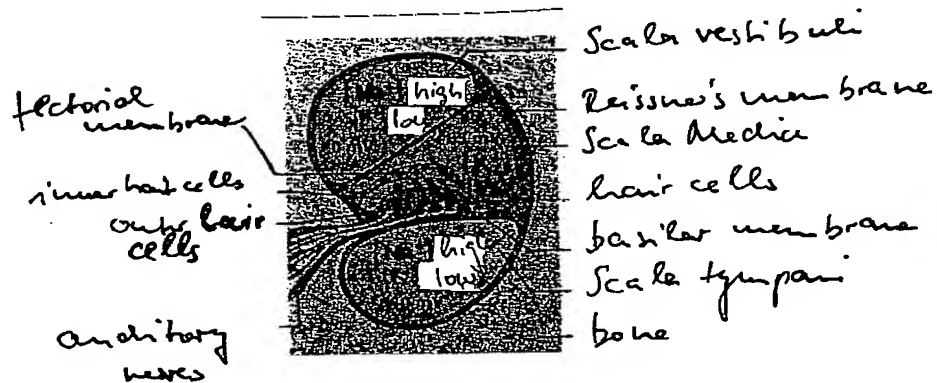


Fig. 20: cross-section of the cochlea

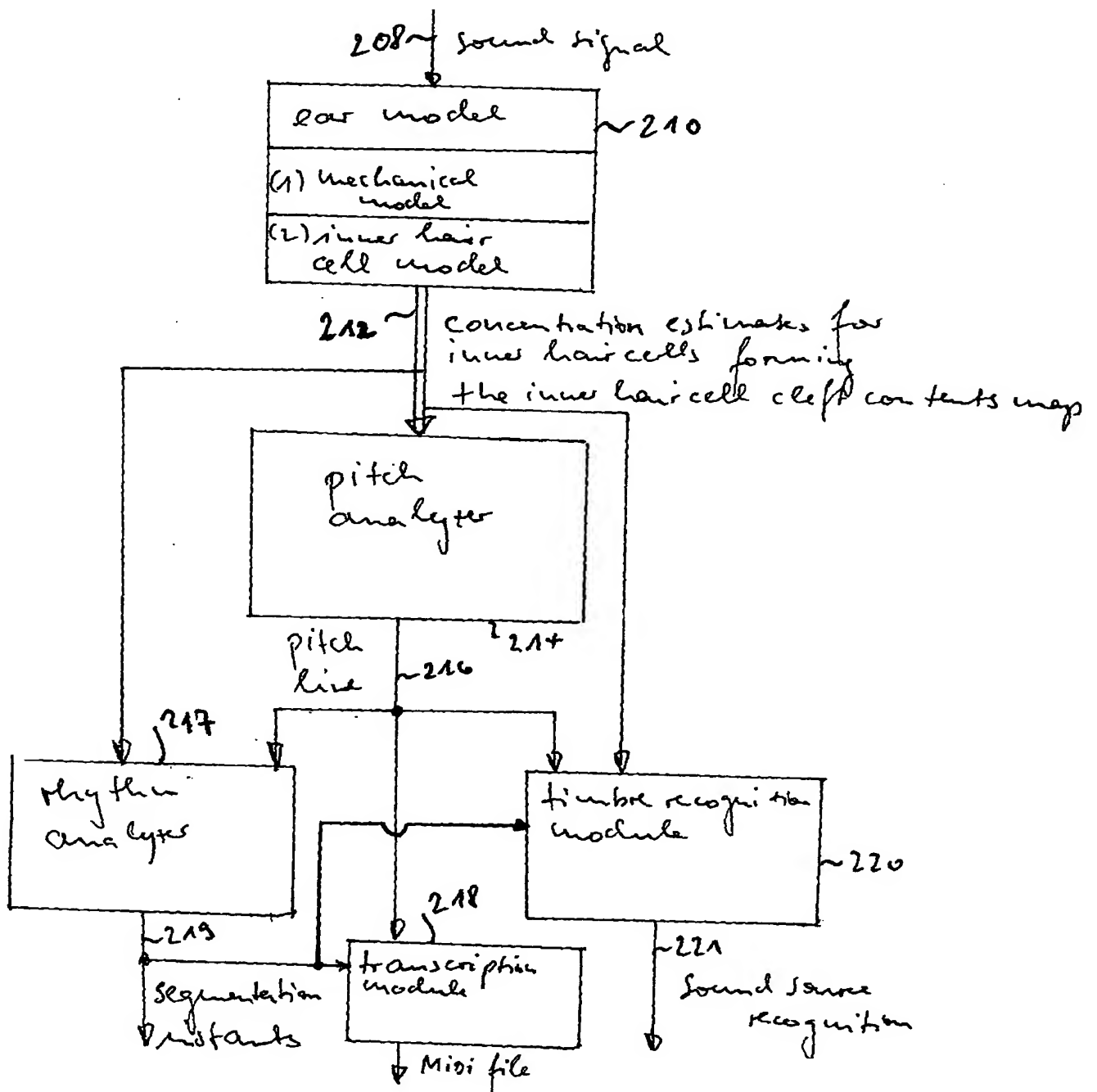


Fig. 21

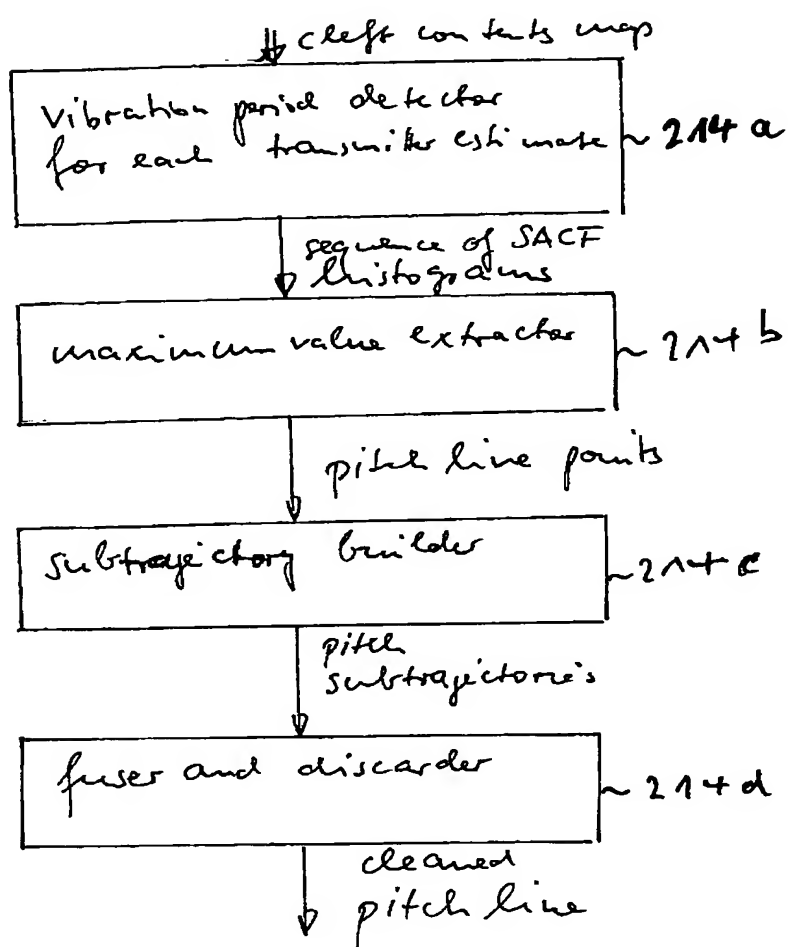


Fig. 22 (pitch analyzer)

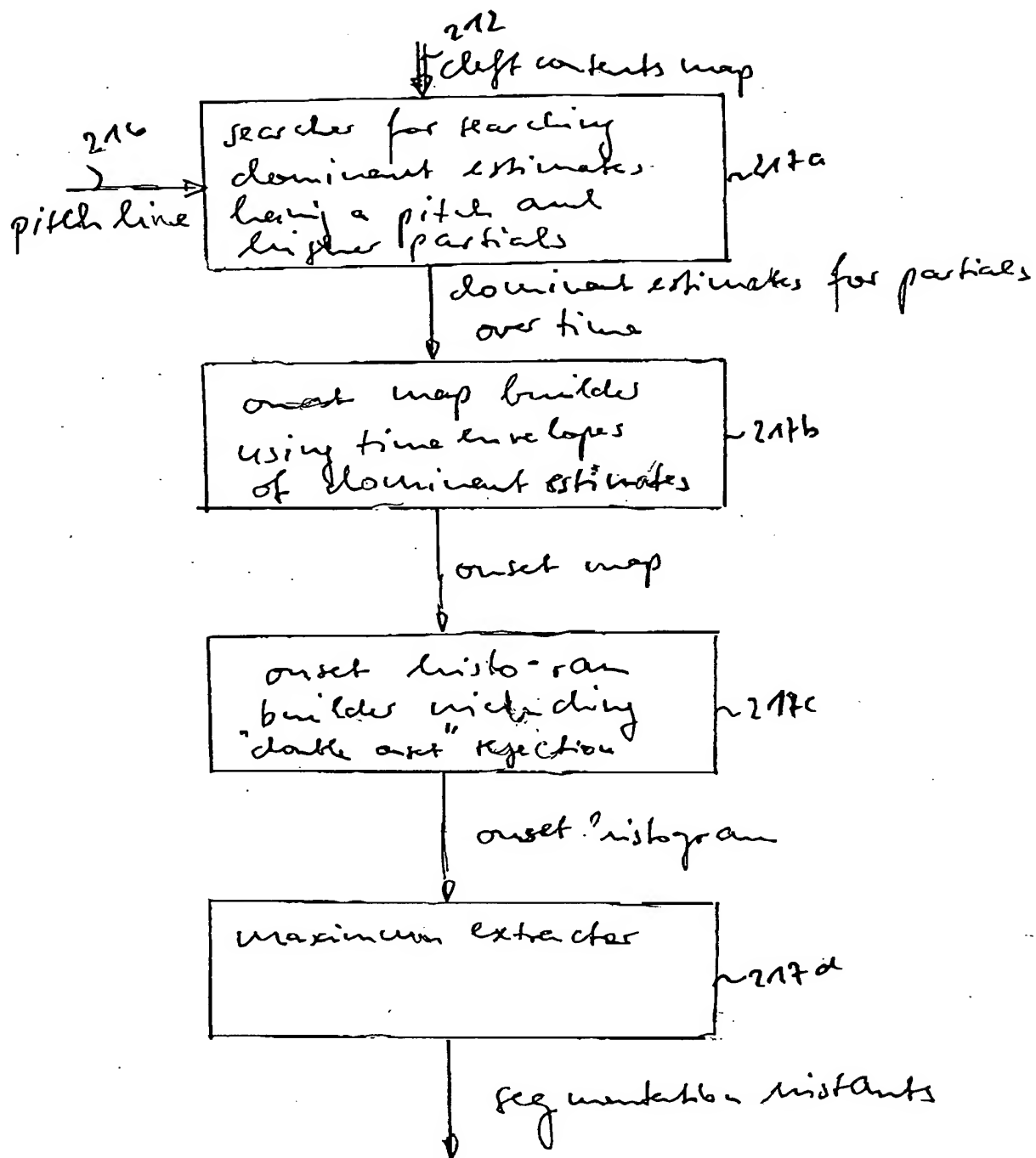


Fig. 23 (rhythm analyzer)